We claim:

1. A method of associating a phenotype with one or more candidate chromosomal regions in a genome of an organism using a phenotypic data structure that represents a difference in a phenotype between different strains of said organism, said genome including a plurality of loci, said method comprising:

establishing a genotypic data structure, said genotypic data structure corresponding to a locus selected from said plurality of loci, said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism;

comparing said phenotypic data structure to said genotypic data structure to form a correlation value; and

repeating said establishing and comparing steps for each locus in said plurality of loci, thereby identifying one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said phenotypic data structure during said comparing step; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation value represent said one or more candidate chromosomal regions.

2. The method of claim 1, wherein an amount of said genome that is included in each locus in said plurality of loci is predetermined.

3. The method of claim 2, wherein said amount is selected from a value in the range of about 0.01 centiMorgans to about 100 centiMorgans.

4. The method of claim 2, wherein said amount is selected from a value in the range of about 5 cM to about 30 cM.

5. The method of claim 1, wherein an instance of said establishing step comprises selecting a locus that is centered on a portion of said genome that is a predetermined distance away from the locus that was selected by a previous instance of said establishing step.

288881.6

6. The method of claim 5, wherein said predetermined distance is measured in centiMorgans.

7. The method of claim 5, wherein said predetermined distance is selected from the range of about 0.0001 centiMorgans to about 30 centiMorgans.

8. The method of claim 5, wherein said predetermined distance is selected from the range of about 2 centiMorgans to about 15 centiMorgans.

9. The method of claim 1, each element in said phenotypic data structure representing a difference in a phenotype between different strains of said organism; wherein, for each element in said phenotypic data structure, said different strains of said organism are selected from a plurality of strains of said organism.

10. The method of claim 9, wherein said difference in said phenotype is determined by a measurement of an attribute corresponding to said phenotype in different strains of said organism.

11. The method of claim 1, each element in said phenotypic data structure representing a difference in said phenotype between a first cluster of strains of said organism and a different second cluster of strains of said organism; wherein, for each element in said phenotypic data structure, said different first and second cluster of strains of said organism are selected from a plurality of clusters of strains of said organism.

12. The method of claim 1, each element in said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism; wherein, for each element in said genotypic data structure, said different strains of said organism are selected from a plurality of strains of said organism.

13. The method of claim 12, wherein an amount that a variation contributes to said at least one component of said locus between different strains of said organism is a function of a distance said variation is away from a center of the locus that corresponds to said genotypic data structure.

47

288881.6

14. The method of claim 13, wherein said genotypic data structure represents a plurality of variations that are distributed about the center of said locus, and said establishing step further comprises:

5   fitting a distribution of said plurality of variations about the center of said locus with a probability function; and

weighting each variation by a corresponding value derived from said probability function such that variations further from the center of said locus are downweighted so that they contribute less to said genotypic data structure than loci

10   that are closer to said center of said locus.

15. The method of claim 14 wherein said probability function is a Gaussian probability distribution, a Poisson distribution, or a Lorentzian distribution.

15   16. The method of claim 1, each element in said genotypic data structure representing a variation of at least one component of said locus between a first cluster of strains of said organism and a different second cluster of strains of said organism; wherein, for each element in said genotypic data structure, said different first and second clusters of strains of said organism are selected from a plurality of strains of said organism.

20

17. The method of claim 1, wherein said correlation value is formed in accordance with the expression:

25    $$c(P, G^L) = \frac{\Sigma^i\, (p(i) - <P>)\, (g(i) - <G^L>)}{\{[\,\Sigma^i\, (p(i) - <P>)^2\,]\, [\Sigma^i\, (g(i) - <G^L>)^2]\}^{1/2}}$$

where,

c(P, G^L) is said correlation value;

30   p(i) is a value of the ith element of said phenotypic data structure;

g(i) is a value of the ith element of said genotypic data structure;

<P> is a mean value of all elements in said phenotypic data structure;

and

<G^L> is a mean value of all elements in said genotypic data structure.

48

18. The method of claim 1, wherein said correlation value is weighted by a number of components in said locus.

5   19. The method of claim 1, wherein each said component is a single nucleotide polymorphism.

20. The method of claim 1, wherein said correlation value is formed in accordance with the expression:

10

$$c(P, G^L) = \frac{[\Sigma^i \, (p(i) - <P>) \, (g(i) - <G^L>)] \times Z}{\{[\Sigma^i \, (p(i) - <P>)^2] \, [\Sigma^i \, (g(i) - <G^L>)^2]\}^{1/2}}$$

15   where,

   $c(P, G^L)$ is said correlation value;

   $p(i)$ is a value of the ith element of said phenotypic data structure;

   $g(i)$ is a value of the ith element of said genotypic data structure;

   $<P>$ is a mean value of all elements in said phenotypic data structure;

20   $<G^L>$ is a mean value of all elements in said genotypic data structure;

   and

   Z is a function of a number of components in said locus having a variation between different strands of said organism.

25   21. The method of claim 20, wherein said function is selected from the group consisting of taking the square root of Z, squaring Z, raising Z by the power of a positive integer, taking a logarithm of Z, and taking an exponential of Z.

22. The method of claim 1, wherein said correlation value is a correlative measure
30   cm that is computed in accordance with the expression:

$$cm(P, G^L) = \frac{[\Sigma^i \, (p(i) - <P>) \, (g(i) - <G^L>)]}{\{[\Sigma^i \, (p(i) - <P>)^2]\}^{1/2}}$$

35

where,

$cmv(P, G^L)$ is said correlative measure;

$p(i)$ is a value of the ith element of said phenotypic data structure;

$g(i)$ is a value of the ith element of said genotypic data structure;

$<P>$ is a mean value of all elements in said phenotypic data structure; and

$<G^L>$ is a mean value of all elements in said genotypic data structure;

23. The method of claim 1, wherein said correlation value is formed using an algorithm selected from the group consisting of regression analysis, regression analysis with data transformations, a Pearson correlation, a Spearman rank correlation, a regression tree and concomitant data reduction, partial least squares, and canonical analysis.

24. The method of claim 1, wherein said repeating step further comprises:

computing (i) a mean correlation value that represents a mean of each said correlation value formed during instances of said comparing step; and (ii) a standard deviation of said mean correlation value based on each said correlation value formed during instances of said comparing step;

wherein, said one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures compared to said phenotypic data structure during said comparing step are identified by selecting genotypic data structures that form a correlation value that is a predetermined number of standard deviations above said mean correlation value.

25. The method of claim 1, wherein each said variation in said genotypic data structure is obtained from a variation in a single nucleotide polymorphism database, a microsatellite marker database, a restriction fragment length polymorphism database, a short tandem repeat database, a sequence length polymorphism database, or an expression profile database.

26. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a

computer program mechanism embedded therein, the computer program mechanism comprising:

a genotypic database for storing variations in genomic sequences of a plurality of strains of an organism;

5   a phenotypic data structure that represents a difference in a phenotype between different strains of said organism; and

a program module for associating a phenotype with one or more candidate chromosomal regions in a genome of said organism, said genome including a plurality of loci, said program module comprising:

10   instructions for establishing a genotypic data structure, said genotypic data structure corresponding to a locus selected from a plurality of loci, said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism stored in said genotypic database;

instructions for comparing said phenotypic data structure to said genotypic data structure to form a correlation value; and

15   instructions for repeating said instructions for establishing and instructions for comparing for each locus in said plurality of loci, thereby identifying one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said phenotypic data structure by said

20   instructions for comparing; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation value represent said one or more candidate chromosomal regions.

27. The computer program product of claim 26, wherein an amount of said genome
25   that is included in each locus in said plurality of loci is predetermined.

28. The computer program product of claim 27, wherein said amount is selected from a value in the range of about 0.01 centiMorgans to about 100 centiMorgans.

30   29. The computer program product of claim 27, wherein said amount is selected from a value in the range of about 5 cM to about 30 cM.

30. The computer program product of claim 26, wherein an instance of said instructions for establishing comprises instructions for selecting a locus that is

51   288881.6

centered on a portion of said genome that is a predetermined distance away from the locus that was selected by a previous instance of said instructions for establishing.

31. The computer program product of claim 30, wherein said predetermined distance is measured in centiMorgans.

32. The computer program product of claim 30, wherein said predetermined distance is selected from the range of about 0.0001 centiMorgans to about 30 centiMorgans.

33. The computer program product of claim 30, wherein said predetermined distance is selected from the range of about 2 centiMorgans to about 15 centiMorgans.

34. The computer program product of claim 26, each element in said phenotypic data structure representing a difference in said phenotype between different strains of said organism; wherein, for each element in said phenotypic data structure, said different strains of said organism are selected from said plurality of strains of said organism represented in said genotypic database.

35. The computer program product of claim 34, wherein said difference in said phenotype is determined by a measurement of an attribute corresponding to said phenotype in said different strains of said organism that are represented in said genotypic database.

36. The computer program product of claim 34, each element in said phenotypic data structure representing a difference in said phenotype between a first cluster of strains of said organism and a different second cluster of strains of said organism; wherein, for each element in said phenotypic data structure, said different first and second cluster of strains of said organism are selected from a plurality of clusters of strains of said organism that are represented in said genotypic database.

37. The computer program product of claim 26, each element in said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism; wherein, for each element in said genotypic data

288881.6

structure, said different strains of said organism are selected from said plurality of strains of said organism represented in said genotypic database.

38. The computer program product of claim 26, wherein an amount that a variation contributes to said at least one component of said locus between different strains of said organism is a function of a distance said variation is away from a center of the locus that corresponds to said genotypic data structure.

39. The computer program product of claim 26, wherein said genotypic data structure represents a plurality of variations that are distributed about the center of said locus, and said instructions for establishing further comprise:

instructions for fitting a distribution of said plurality of variations about the center of said locus with a probability function; and

instructions for weighting each variation by a corresponding value derived from said probability function such that variations further from the center of said locus are downweighted so that they contribute less to said genotypic data structure than loci that are closer to said center of said corresponding locus.

40. The computer program product of claim 39 wherein said probability function is a Gaussian probability distribution, a Poisson distribution, or a Lorentzian distribution.

41. The computer program product of claim 26, each element in said genotypic data structure representing a variation of at least one component of said locus between a first cluster of strains of said organism and a different second cluster of strains of said organism; wherein, for each element in said genotypic data structure, said different first and second clusters of strains of said organisms are selected from said plurality of strains of said organism represented in said genotypic database.

42. The computer program product of claim 26 wherein said instructions for comparing include instructions for forming said correlation value in accordance with the expression:

$$c(P, G^L) = \frac{\sum^i (p(i) - <P>) \; (g(i) - <G^L>)}{\{[\sum^i (p(i) - <P>)^2] [\sum^i (g(i) - <G^L>)^2]\}^{1/2}}$$

where,

c(P, G^L) is said correlation value;

p(i) is a value of the ith element of said phenotypic data structure;

g(i) is a value of the ith element of said genotypic data structure;

<P> is a mean value of all elements in said phenotypic data structure;

and

<G^L> is a mean value of all elements in said genotypic data structure.

43. The computer program product of claim 26, wherein said correlation value is weighted by a number of components in said locus.

44. The computer program product of claim 26, wherein each said component is a single nucleotide polymorphism.

45. The computer program product of claim 26, wherein said instructions for comparing include instructions for forming said correlation value in accordance with the expression:

$$c(P, G^L) = \frac{[\sum^i (p(i) - <P>) \; (g(i) - <G^L>)] \times Z}{\{[\sum^i (p(i) - <P>)^2] [\sum^i (g(i) - <G^L>)^2]\}^{1/2}}$$

where,

c(P, G^L) is said correlation value;

p(i) is a value of the ith element of said phenotypic data structure;

g(i) is a value of the ith element of said genotypic data structure;

<P> is a mean value of all elements in said phenotypic data structure;

<G^L> is a mean value of all elements in said genotypic data structure;

and

Z is a function of a number of components in said locus having a variation between different strands of said organism.

54

288881.6

46. The computer program product of claim 43, wherein said function is selected from the group consisting of taking the square root of Z, squaring Z, raising Z by the power of a positive integer, taking a logarithm of Z, and taking an exponential of Z.

47. The computer program product of claim 26, wherein said instructions for comparing include instructions for forming said correlation value in accordance with a correlative measure cm that is computed in accordance with the expression:

$$cm(P, G^L) = \frac{[\sum^i (p(i) - <P>) \ (g(i) - <G^L>)]}{\{[\sum^i (p(i) - <P>)^2]\}^{1/2}}$$

where,

$cm(P, G^L)$ is said correlative measure;

$p(i)$ is a value of the ith element of said phenotypic data structure;

$g(i)$ is a value of the ith element of said genotypic data structure;

$<P>$ is a mean value of all elements in said phenotypic data structure;

and

$<G^L>$ is a mean value of all elements in said genotypic data structure.

48. The computer program product of claim 26, wherein said instructions for comparing include instructions for forming said correlation value by an algorithm selected from the group consisting of regression analysis, regression analysis with data transformations, a Pearson correlation, a Spearman rank correlation, a regression tree and concomitant data reduction, partial least squares, and canonical analysis.

49. The computer program product of claim 26, wherein said instructions for repeating further comprise:

instructions for computing (i) a mean correlation value that represents a mean of each said correlation value formed during instances of said instructions for comparing; and (ii) a standard deviation of said mean correlation value based on each said correlation value formed during instances of said instructions for comparing;

wherein, said one or more genotypic data structures that form a high

correlation value relative to all other genotypic data structures compared to said

phenotypic data structure by said instructions for comparing are identified by

selecting genotypic data structures that form a correlation value that is a

5   predetermined number of standard deviations above said mean correlation value.

50.   The computer program product of claim 26, wherein said genotypic database is a

single nucleotide polymorphism database, a microsatellite marker database, a

restriction fragment length polymorphism database, a short tandem repeat database, a

10   sequence length polymorphism database, an expression profile database, or a DNA

methylation database; and said variation in said genotypic data structure is obtained

from said genotypic database.

51.   A computer program product for use in conjunction with a computer system, the

15   computer program product comprising a computer readable storage medium and a

computer program mechanism embedded therein, the computer program mechanism

comprising:

       a genotypic database for storing variations in genomic sequences of a plurality

of strains of an organism;

20   a phenotypic data structure, each element in said phenotypic data structure

representing a difference in said phenotype between different strains of said organism;

and

       a program module for associating a phenotype with one or more candidate

chromosomal regions in a genome of said organism, said genome including a plurality

25   of loci, said program module comprising:

       instructions for identifying a genotypic data structure, said genotypic data

structure corresponding to a locus selected from said plurality of loci, each element in

said genotypic data structure representing a variation of at least one component of

said locus between different strains of said organism;

30   instructions for comparing said phenotypic data structure to said genotypic

data structure to form a correlation value; and

       instructions for repeating said instructions for identifying and said instructions

for comparing, for each locus in said plurality of loci, thereby identifying one or more

genotypic data structures that form a high correlation value relative to all other

56

genotypic data structures that are compared to said phenotypic data structure by said instructions for comparing; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation value represent said one or more candidate chromosomal regions.

52. A computer system for associating a phenotype with one or more candidate chromosomal regions in a genome of an organism, said genome including a plurality of loci, the computer system comprising:

a central processing unit;

a memory, coupled to the central processing unit, the memory storing:

a genotypic database for storing variations in genomic sequences of a plurality of strains of said organism;

a phenotypic data structure that represents a difference in a phenotype between different strains of said organism; and

a program module, said program module comprising:

instructions for establishing a genotypic data structure, said genotypic data structure corresponding to a locus selected from a plurality of loci, said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism stored in said genotypic database;

instructions for comparing said phenotypic data structure to said genotypic data structure to form a correlation value; and

instructions for repeating said instructions for establishing and said instructions for comparing, for each locus in said plurality of loci, thereby identifying one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said phenotypic data structure by said instructions for comparing; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation value represent said one or more candidate chromosomal regions.

53. The computer system of claim 52, each element in said phenotypic data structure representing a variation in said phenotype between different strains of said organism; wherein, for each element in said phenotypic data structure, said different strains of said organism are selected from said plurality of strains of said organism represented in said genotypic database.

54. The computer system of claim 53, wherein said difference in a phenotype is determined by a measurement of an attribute corresponding to said phenotype in said different strains of said organism that are represented in said genotypic database.

55. The computer system of claim 52, each element in said phenotypic data structure representing a variation in said phenotype between a first cluster of strains of said organism and a different second cluster of strains of said organism; wherein, for each element in said phenotypic data structure, said different first and second cluster of strains of said organism are selected from a plurality of clusters of strains of said organism that are represented in said genotypic database.

56. The computer system of claim 52, each element in said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism; wherein, for each element in said genotypic data structure, said different strains of said organism are selected from said plurality of strains of said organism represented in said genotypic database.

57. The computer system of claim 52, each element in said genotypic data structure representing a variation of at least one component of said locus between a first cluster of strains of said organism and a different second cluster of strains of said organism; wherein, for each element in said genotypic data structure, said different first and second clusters of strains of said organisms are selected from said plurality of strains of said organism represented in said genotypic database.

58. The computer system of claim 52, wherein said instructions for comparing include instructions for forming said correlation value in accordance with the expression:

$$c(P, G^L) = \frac{\Sigma^i \, (p(i) - <P>) \; (g(i) - <G^L>)}{\{[\, \Sigma^i \, (p(i) \; - <P>)^2] \, [\Sigma^i \, (g(i) \; - <G^L>)^{\,2}]\}^{1/2}}$$

where,

$c(P, G^L)$ is said correlation value;

p(i) is a value of the ith element of said phenotypic data structure;

g(i) is a value of the ith element of said genotypic data structure;

$<P>$ is a mean value of all elements in said phenotypic data structure; and

$<G^L>$ is a mean value of all elements in said genotypic data structure.

59. The computer system of claim 52, wherein said instructions for comparing include instructions for forming said correlation value by an algorithm selected from the group consisting of regression analysis, regression analysis with data transformations, a Pearson correlation, a Spearman rank correlation, a regression tree and concomitant data reduction, partial least squares, and canonical analysis.

60. The computer system of claim 52, wherein said instructions for repeating further comprise:

instructions for computing (i) a mean correlation value that represents a mean of each said correlation value formed during instances of said instructions for comparing; and (ii) a standard deviation of said mean correlation value based on each said correlation value formed during instances of said instructions for comparing;

wherein, said one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures compared to said phenotypic data structure by said instructions for comparing are identified by selecting genotypic data structures that form a correlation value that is a predetermined number of standard deviations above said mean correlation value.

61. The computer system of claim 52, wherein said genotypic database is a single nucleotide polymorphism database, a microsatellite marker database, a restriction fragment length polymorphism database, a short tandem repeat database, a sequence length polymorphism database, an expression profile database, or a DNA methylation database; and said variation in said genotypic data structure is obtained from said genotypic database.

62. A method of associating a phenotype with one or more candidate chromosomal regions in a genome of an organism using a phenotypic data structure that represents

alterations in phenotypes between different strains in a plurality of strains of said organism,

said phenotypic data structure including a description of each said alteration and individual elements of said phenotypic data structure including an amount of

5   alteration between different strains of said organism selected from said plurality of strains of said organism,

said genome including a plurality of loci, each said loci representing one or more positions within said genome,

said method comprising:

10   establishing a unique individual variation matrix for each said one or more positions represented by said loci, wherein an element within each said unique individual variation matrix represents an allelic comparison between different strains of said organism that are selected from said plurality of strains of said organism;

summing corresponding elements in each said unique individual matrix to

15   form a genotypic data structure;

comparing said phenotypic data structure to said genotypic data structure to form a correlation value; and

repeating said establishing, summing and comparing steps, for each locus in said plurality of loci, thereby identifying one or more genotypic data structures that

20   form a high correlation value relative to all other genotypic data structures that are compared to said phenotypic data structure during said comparing step; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation value represent said one or more candidate chromosomal regions associated with said phenotype.

25

63. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

30   a genotypic database for storing variations in genomic sequences of a plurality of strains of an organism;

a phenotypic data structure that represents alterations in phenotypes between different strains of said organism selected from said plurality of strains of said organism, said phenotypic data structure including a description of each said

alteration and individual elements of said phenotypic data structure including an amount of alteration between different strains in said plurality of strains of said organism; and

a program module for associating a phenotype with one or more candidate chromosomal regions in a genome of said organism, said genome including a plurality of loci, each said loci representing one or more positions within said genome, said program module comprising:

instructions for establishing a unique individual variation matrix for each said one or more positions represented by said loci, wherein an element within each said unique individual variation matrix represents an allelic comparison of values stored in said genotypic database between different strains of said organism that are selected from said plurality of strains of said organism;

instructions for summing corresponding elements in each said unique individual matrix to form a genotypic data structure;

instructions for comparing said phenotypic data structure to said genotypic data structure to form a correlation value; and

instructions for repeating said instructions for establishing, summing and comparing, for each locus in said plurality of loci, thereby identifying one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said phenotypic data structure during said comparing step; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation value represent said one or more candidate chromosomal regions associated with said phenotype.

64. A computer system for associating a phenotype with one or more candidate chromosomal regions in a genome of an organism, said genome including a plurality of loci, each said loci representing one or more positions within said genome, said program module comprising:

a central processing unit;

a memory, coupled to the central processing unit, the memory storing:

a genotypic database for storing variations in genomic sequences of a plurality of strains of said organism;

a phenotypic data structure that represents alterations in phenotypes between different strains in said plurality of strains of said organism, said phenotypic data

structure including a description of each said alteration and individual elements of said phenotypic data structure including an amount of alteration between different strains in said plurality of strains of said organism; and

a program module, said program module comprising:

5   instructions for establishing a unique individual variation matrix for each said one or more positions represented by said loci, wherein an element within each said unique individual variation matrix represents an allelic comparison of values stored in said genotypic database between different strains of said organism that are selected from said plurality of strains of said organism;

10   instructions for summing corresponding elements in each said unique individual matrix to form a genotypic data structure;

instructions for comparing said phenotypic data structure to said genotypic data structure to form a correlation value; and

instructions for repeating said instructions for establishing, summing and 15   comparing, for each locus in said plurality of loci, thereby identifying one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said phenotypic data structure during said comparing step; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation represent said one or more candidate 20   chromosomal regions associated with said phenotype.

65. A method of determining a portion of a genome of an organism that is responsive to a perturbation, the method comprising:

producing a first phenotypic data structure that represents a difference in a first 25   phenotype between different strains of said organism, said genome including a plurality of loci, wherein said first phenotype is measured for each said different strain of said organism when each said different strain is in a first state;

establishing a genotypic data structure, said genotypic data structure corresponding to a locus selected from said plurality of loci, said genotypic data 30   structure representing a variation of at least one component of said locus between different strains of said organism;

comparing said first phenotypic data structure to said genotypic data structure to form a correlation value;

288881.6

repeating said establishing and comparing steps for each locus in said plurality of loci, thereby identifying a first set of genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said first phenotypic data structure during said comparing step;

5   computing a second phenotypic data structure that represents a difference in a second phenotype between different strains of said organism, wherein said second phenotype is measured for each said different strain of said organism when each said different strain is in a second state that is produced by exposing each said different strain of said organism to a perturbation;

10   correlating said second phenotypic data structure to said genotypic data structure to form a correlation value;

repeating said computing and correlating steps for each locus in said plurality of loci, thereby identifying a second set of genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to

15   said second phenotypic data structure during said correlating step; and

resolving a dissimilarity in said first set of genotypic data structures and said second set of genotypic structures, thereby determining said portion of said genome of said organism that is responsive to said perturbation.

20   66. The method of claim 65 wherein said perturbation is a pharmacological agent.

67. The method of claim 65 wherein said perturbation is a chemical compound having a molecular weight of less than 1000 Daltons.

25   68. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

a program module for determining a portion of a genome of an organism that

30   is responsive to a perturbation, the method comprising:

instructions for producing a first phenotypic data structure that represents a difference in a first phenotype between different strains of said organism, said genome including a plurality of loci, wherein said first phenotype is measured for

each said different strain of said organism when each said different strain is in a first state;

instructions for establishing a genotypic data structure, said genotypic data structure corresponding to a locus selected from said plurality of loci, said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism;

instructions for comparing said first phenotypic data structure to said genotypic data structure to form a correlation value;

instructions for repeating said instructions for establishing and said instructions for comparing for each locus in said plurality of loci, thereby identifying a first set of genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said first phenotypic data structure during said comparing step;

instructions for computing a second phenotypic data structure that represents a difference in a second phenotype between different strains of said organism, wherein said second phenotype is measured for each said different strain of said organism when each said different strain is in a second state that is produced by exposing each said different strain of said organism to a perturbation;

instructions for correlating said second phenotypic data structure to said genotypic data structure to form a correlation value;

instructions for repeating said computing and correlating steps for each locus in said plurality of loci, thereby identifying a second set of genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said second phenotypic data structure during said correlating step; and

instructions for resolving a dissimilarity in said first set of genotypic data structures and said second set of genotypic structures, thereby determining said portion of said genome of said organism that is responsive to said perturbation.

69. The computer program product of claim 68 wherein said perturbation is a pharmacological agent.

70. The computer program product of claim 68 wherein said perturbation is a chemical compound having a molecular weight of less than 1000 Daltons.

288881.6

71. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

5   a program module for associating a phenotype with one or more candidate chromosomal regions in a genome of said organism, said genome including a plurality of loci, said program module comprising:

        instructions for accessing a genotypic data structure, said genotypic data structure corresponding to a locus selected from a plurality of loci, said genotypic data structure representing a variation of at least one component of said locus between different strains of said organism stored in a genotypic database;

        instructions for comparing a phenotypic data structure to said genotypic data structure to form a correlation value; and

        instructions for repeating said instructions for establishing and instructions for comparing for each locus in said plurality of loci, thereby identifying one or more genotypic data structures that form a high correlation value relative to all other genotypic data structures that are compared to said phenotypic data structure by said instructions for comparing; wherein the loci that correspond to said one or more genotypic data structures that form a high correlation value represent said one or more candidate chromosomal regions.